ANALYSIS, SECURE ACCESS TO, AND TRANSMISSION OF ARRAY IMAGES

5 Related Application

Priority is hereby claimed to U.S. Provisional Application No. 60/426839, filed 11/15/2002.

Background

Recent rapid advances in molecular biology have created more demand for 10 high volume testing based on the need to screen ever larger compound libraries, validate ever increasing numbers of genetic markers and test ever more diversified patient populations. This has led to the development of new array formats, particularly for nucleic acid and protein-protein interaction analysis, which increase parallel processing by performing requisite assays in a "multiplexed" format.

15 Conventionally, such assays are performed by producing arrays of nucleic acids and antibodies by way of "spotting" or "printing" of aliquot solutions on filter paper, blotting paper or other substrates. However, notwithstanding their widespread current use in academic research targeting gene expression and protein profiling, arrays produced by spotting have shortcomings, particularly in applications placing 20 high demands on accuracy and reliability and where large sample volume and high throughput is required. In another more recently developed technique, spatially encoded probe arrays are produced by way of in-situ photochemical oligonucleotide synthesis. However, this technology is limited in practice to producing short oligonucleotide probes – and requiring alternative technologies for the production of

1

cDNA and protein arrays - and precludes rapid probe array customization given the time and cost involved in the requisite redesign of the photochemical synthesis process.

In addition to these inherent difficulties in assay performance, spatially encoded arrays produced by methods of the art generally produce data of such poor quality that specialized scanners are required to extract data of useable quality. Commercial systems available for this purpose require confocal laser scanning - a slow process which must be repeated for each desired signal color - and limit the spatial resolution to ~5 μm.

In order to resolve many of the problems associated with diagnostic and analytical uses of "spotted arrays" of oligonucleotides and proteins (as outlined in "Multianalyte Molecular Analysis Using Application-Specific Random Particle Arrays," U.S. Application Serial No. 10/204,799, filed on 8/23/2002; WO 01/98765), arrays of oligonucleotides or proteins arrays can be formed by displaying these capture moieties on chemically encoded microparticles ("beads") which are then assembled into planar arrays composed of such encoded functionalized carriers. *See* U.S. Patent Application Serial No. 10/271,602 "Multiplexed Analysis of Polymorphic Loci by Concurrent Interrogation and Enzyme-Mediated Detection," filed 10/15/2002, and Serial No. 10/204,799 *supra.*

Microparticle arrays displaying oligonucleotides or proteins of interest can be assembled by light-controlled electrokinetic assembly near semiconductor surfaces (*see, e.g.,* U.S. Patent Nos. 6,468,811; 6,514,771; 6,251,691) or by a direct

disposition assembly method (previously described in Provisional Application Serial No. 60/343,621, filed December 28, 2001 and in U.S. Application Serial No. 10/192,352, filed July 9, 2002).

To perform nucleic acid or protein analysis, such encoded carrier arrays are placed in contact with samples anticipated to contain target polynucleotides or protein ligands of interest. Capture of target or ligand to particular capture agents displayed on carriers of corresponding type as identified by a color code produces, either directly or indirectly by way of subsequent decoration, in accordance with one of several known methods, an optical signature such as a fluorescence signal. The identity of capture agents including probes or protein receptors (referred to herein sometimes also collectively as "receptors") generating a positive assay signal can be determined by decoding carriers within the array.

These microparticle arrays can exhibit a number of spectrally distinguishable types of beads within an area small enough to be viewed in a microscope field. It is possible to achieve a high rate of image acquisition because the arrays obviate the need for confocal laser scanning (as used with spotted or *in-situ* synthesized arrays) and instead permit the use of direct ("snapshot") multicolor imaging of the entire array under a microscope. If the system could be automated further, such that, for example, the microscope is automatically repositioned to optimally capture images from multiple arrays present on a multichip carrier and to positions optimizing decoding of the array, this would facilitate unattended acquisition of large data lots from multiplexed assays.

In one format using microbead arrays, the encoding capacity of a chip (which

3

includes several distinct subarrays) can be increased even where using the same set of color codes for the beads in each subarray. When subarrays are spatially distinct, the encoding capacity becomes the product of the number of bead colors and the number of subarrays.

In order to match the rates of data acquisition enabled by direct imaging, rapid and robust methods of image processing and analysis are required to extract quantitative data and to produce encrypted and compact representations suitable for rapid transmission, particularly where there is off-site analysis and data storage. Transmission of data should be secure, and should be accessible only by authorized parties, including the patient but, because of privacy concerns, not to others.

Summary

Disclosed are methods of increasing the confidence of the analysis, and for rapid and automated decoding of encoded arrays used in assays, assay data recorded in the form of images generated from arrays of ligand-receptor interactions; and more particularly, where different receptors are associated with different encoded microparticles ("beads"), and results are determined upon decoding of the arrays. Also disclosed are methods for transmitting and archiving data from such assay arrays in a manner such that access is limited to authorized persons, and such that the chance of assigning one patient's results to another are minimized. These methods are particularly useful where assays are performed at decentralized user ("client") sites, because the methods permit secure exchange of data between the client and a central facility ("information keeper"), where the data can be centrally decoded and

4

analyzed so as to provide greater reliability, and then archived in a restricted manner where only authorized users have access.

In a centralized regime, patient samples, collected in the field, are sent for analysis to a central location, where they are assayed. Results are provided to

5   authorized users by remote transmission. Users, while relieved of any responsibility relating to assay completion and data analysis, are faced with the loss of control over the assay implementation and analysis and may face the inconvenience of significant delay. Non-standard assays may be unavailable or prohibitively expensive. In addition, this service ordinarily will not be suitable for perishable samples, or large

10   collections of samples, such as those created in a pharmaceutical research laboratory.

In a decentralized paradigm, analytical instrumentation, such as microplate readers complete with all requisite software, are distributed to users who perform assays, record results, and may also perform subsequent data analysis. Alternatively, assay results may be transmitted to a central facility for decoding, analysis,

15   processing and archiving, and such centralized procedures may provide greater reliability.

The analysis server model (useful, *inter alia*, for molecular diagnostics), as disclosed herein, expands upon these paradigms by combining decentralization of assay performance with centralized data analysis. That is, while assay performance

20   and data generation are at user facilities, critical aspects of subsequent data analysis and related services may be performed in a centralized location which is accessible to authorized users in a two-way mode of communication via public or private computer networks. The analysis server model can be applied to assays performed in

a highly parallel array format requiring only a simple imaging instrument, such as a microscope, to record complex assay data, but requiring advanced methods of analysis and mathematical modeling to reliably process and analyze assay data. Images, recorded at a user location, are uploaded to a centralized location where such analysis are performed, results being made available to authorized users in real time.

The methods and processes are further explained below with reference to the drawings.

Brief Description of the Drawings

Fig. 1 displays a 1x8 multichip carrier, where chips with encoded arrays on a surface are housed in the carrier's wells.

Fig. 2 is a flowchart of an imaging system, including image acquisition, a decoder and a reader for image processing, and an analyzer for data analysis.

Fig. 3 is a flowchart illustrating an exemplary system where assay data can be decoded and centrally analyzed.

Fig. 4 is a flowchart illustrating viewing a multiplexed assay and recording of the assay image.

Fig. 5 is a flowchart depicting sample collection, performing of assays and acquisition of an assay image.

Fig. 6 is a flowchart of the formation of a decoding image and decoding data records where the resulting information can be stored in a database.

Fig. 7 is a diagram illustrating the, reading and processing of an assay image and recording of an assay data record, where the resulting information can be stored in a database.

Fig. 8 is a flowchart illustrating the recording, analysis and decoding of assay data, and combining with an assay data record to generate a decoded assay data record, where the resulting information can be stored in a database.

Fig. 9A shows a grid (with extra rows and columns) in unshifted position (0,0).

Fig. 9B shows the grid of Fig. 9A and in a configuration shifted to position (1,1).

Fig. 9C illustrates the seven possible shifting positions for the grid of 9A and 9B.

Fig. 10A is an illustration of a two-party, multiple-user, data exchange.

Fig. 10B illustrates a three-party data exchange, where one party is the patient.

Fig. 10C illustrates more specifically a secure data transaction between three parties, where one party is the patient.

Figs. 11A to 11D depict the DECODER Graphical User Interface (GUI) illustrating steps in the course of processing a decoding image.

Fig. 12 illustrates a constructed Decoding Map.

Fig. 13 shows a READER Graphical User Interface (GUI) illustrating steps in the course of processing a assay image.

Fig. 14 illustrates a decoded assay data record in histogram form.

Fig. 15A illustrates mesh overall orientation with the shortest path algorithm, wherein L_h^{ref} and L_v^{ref} are the reference lines. By applying shortest path, L_h and L_v are found. It can be seen that when L_h is shorter than L_v, then the overall orientation of the mesh is 0 degrees.

Fig. 15B illustrates vertical line grid partition by the shortest path algorithm.

7

Detailed Description

Fig. 1 depicts a multichip carrier (100) displaying a barcode (110) along with a set of 2x2 chips (120) in each of eight positions, each such chip comprising one or more random array(s) of encoded microparticles (130). The decoding of the microparticles 130 and analysis and transmission of the data can be performed in accordance with the methods described herein. Each of the 2x2 chip subarrays (*i.e.*, four chips) correspond with one individual patient sample. Thus, given a set of 128 distinguishable colors for the beads in the array, this can produce an array complexity of 4 × 128 = 512 for each individual. The 1x 8 sample format shown in Fig. 1 with a center-to-center distance of 9 mm matches the standard 8x12 microwell format and is readily scaled to nx8 for high-throughput sample handling (as described further in U.S. Patent Application Serial No. 10/192,352).

As illustrated in Fig.2, the imaging system comprises hardware for acquiring a decoding image (200), means for acquiring an assay image (210), a decoder (220), a reader (230), and an analyzer (240). Acquiring and recording a decoding image or an assay image can be accomplished using the system shown in Fig. 4.

Molecular interaction analysis on a random array of encoded microparticles or beads, where each subpopulation of such encoded beads displays a unique receptor molecule, for example, an oligonucleotide, or protein molecule, can be performed using these systems. Each such receptor is designed so as to be able to form a molecular complex with a cognate target analyte or ligand, the formation of a complex resulting in an optical signature (for example, the ligand can be

8

fluorescently labeled and thus detectable following complex formation) that is detected by acquiring and analyzing images of the array. Assay results take the form of an image comprising a set of intensities, each of which is uniquely associated with one receptor-ligand interaction in the array. Such assays are well-suited for DNA and protein analysis performed in a multiplexed or parallel mode using an array format, including applications such as genetic expression profiling, polymorphism and mutation analysis, protein-protein analysis including antibody-antigen interaction analysis, organic compound-receptor interactions, and further including all those disclosed in PCT/US01/20179, US Patent Nos. 6,251,691; 6,387,707 and U.S. Patent Application Serial No. 10/271,602, all of which are incorporated herein by reference.

The decoding of encoded arrays and the analysis, interpretation or storage of assay results requiring access to methods and algorithms or access to databases which are not available at the location housing the analytical instrumentation, but are instead available at a remote location, is addressed herein. Special consideration is given to situations in which these operations, or parts thereof such as the collection and preparation of a sample, the actual assay or assays of interest and additional operations such as interpretation or consultation are performed in separate locations. In one embodiment, the methods and apparatus described herein permit complex multianalyte analysis to be performed in locations visited by a patient whose sample is collected and analyzed on site, while providing transaction protocols to perform data analysis and optional additional services, such as data interpretation and archiving to be performed off-site. The methods herein permit the secure exchange

of information recorded on site and analyzed at the site of an application service provider.

Accordingly, the systems and methods provided herein for the analysis of clinical samples or other analytes provide for communication among three (or optionally more) participants, including: (i) the sample originator, for example a patient seeking clinical diagnosis or a biomedical research laboratory seeking analytical services; (ii) the analysis provider (which may or may not also be the tester performing the assay); and (iii) the tester, which performs the assay but is generally provided minimal information about assay outcomes, decoding, or sample origination; and (iv) an optional intermediary, for example a sample collector and/or processor or communicator of results, such as a counselor.

These methods allow the rapid assaying and analysis of customized random encoded bead arrays, where a multiplexed assays are performed on patient samples, and multiple assays may be conducted. Suitable panels may include, for example, a tumor marker panel including antigens such as PSA and other suitable tumor markers, an allergy panel, a pregnancy panel comprising tests for human chorionic gonadotropin, hepatitis B surface antigen, rubella virus, alpha fetoprotein, 3' estradiol and other substances of interest for monitoring a pregnant individual; a hormone panel, an autoimmune disease panel including tests for rheumatoid factors and panel reactive antibodies and other markers associated with autoimmune disorders, a panel of blood-borne viruses and a therapeutic drug panel comprising tests for Cyclosporin, Digoxin and other therapeutic drugs of interest.

In addition, such panels also may include, for example, oligonucleotide

10

probes designed for nucleic acid analysis including analysis of cDNA panels for gene expression profiling, oligonucleotide probe panels designed for the multiplexed analysis of mutations causing genetic diseases such as cystic fibrosis, Tay-Sachs disease, Ashkenazi Jewish diseases including Gaucher disease and others, the analysis of polymorphisms such as those in the Human Leukocyte Antigen complex which determine the degree of compatibility between donor and recipient in transplantation of bone marrow or solid organs, a blood antigen panel for blood typing, the analysis of chromosomal aberrations such as those underlying Down Syndrome and others surveyed in prenatal screening or certain blood-borne cancers such as certain leukemias. The multiplexed nucleic acid analysis involved in assaying of these panels can be performed using either hybridization-mediated detection or hybridization-mediated elongation-mediated detection, as described in U.S. Patent Application Serial No. 10/271,602, entitled: "Multiplexed Analysis of Polymorphic Loci by Concurrent Interrogation and Enzyme-Mediated Detection" filed 10/15/2002.

In either hybridization-mediated detection or hybridization-mediated elongation-mediated detection, an association of polynucleotide in the sample with a probe oligonucleotide on a bead results in an assay signal in the form of an optical signature. For example, in the READ™ format, each encoded bead within the array (where each bead has multiple probes attached thereto) may produce one or more of such optical signatures which are able to be recorded by the systems of the invention. The optical signature can be a fluorescence signature. Optical signatures of interest include, without limitation, luminescence including bioluminescence,

11

chemiluminescencve and electrochemiluminescence. Direct visual signatures resulting, for example, from the transformation of the assay locus, for example, by agglutination of multiple beads, or the attachment of marker particles to assay loci, can also be recorded and analyzed using the methods set forth herein.

5   I.   **Automated High-Throughput Array Imaging for Molecular Interaction Analysis**

I.1   Assay and Decoding Data Records

Fig. 3 is an illustration of the methods to conduct the analysis of analytes using bead arrays assembled on substrates, according to the READ process of multiplexed analysis. The method of Fig. 3 involves producing a bead array (300), obtaining a decoding image (310), processing the decoding image using a decoder (320), and obtaining a decoding data record (330). In parallel, an assay is performed using the bead array (340) to obtain an assay image (350), where the assay image is processed using a reader (321) and the assay data is recorded (360). The decoding data and assay data are then combined and the image is analyzed (370). The decoding image may represent a combination of several distinct images recorded in separate spectral bands or color channels, where the encoding is with chemically and/or physically distinguishable characteristics that uniquely identify the binding agent displayed on the bead surface. For example, when the uniquely distinguishable characteristic is color, a "decoding image" of the bead array (where the bead are immobilized on the substrate) is recorded to reference the color code of constituent beads, where the color code uniquely corresponds to the chemical identity of the binding agents displayed on each individual bead's surface.

12

Referring to Fig. 3, the decoding image (310) may be generated following completion of array assembly by the array manufacturer at the manufacturing site or it may be generated by the user of the array in connection with the completion of a bioassay or other chemical test, either prior to, or subsequent to completion of the assay or test at the test site. The decoding image (310) becomes part of the decoding data record (330), which also contains a variety of identifiers for reagent, microparticle and substrate batch and lot numbers. The decoding image (310) can be generated following array assembly at the manufacturing site and processed to create a decoding data record (330) which is stored in a database on a central server. Access to the decoding data record, or parts thereof, can be accessible in the form of a copy of the record, for example, of copies on a recording medium such as a CD that is distributed along with arrays and multichip carriers or cartridges. Alternatively, access to the decoding data record, or parts thereof, is made accessible in the form of authenticated database accessible only to authorized users as disclosed herein.

Following completion of an assay (340) at a user site, the assay image is recorded (350) and an Assay Data Record (360) is created which serves to record the optical signature(s) indicating the binding of ligand molecules to immobilized receptors. For example, when fluorescence is selected to provide the optical signature of interest, the fluorescence intensity recorded from each position within the array indicates the amount of complex formed in that location by receptor and ligand binding or hybridization. Multiple modes of generating such optical signatures include the direct or indirect labeling of target analytes (for example, by using fluorescent primers to conduct PCR of genomic regions to be assayed) or the

13

introduction of fluorescence by way of probe elongation using labeled nucleotides. See, e.g., U.S. Patent Application Serial No. 10/271,602 "Multiplexed Analysis of Polymorphic Loci by Concurrent Interrogation and Enzyme-Mediated Detection." The assay image forms the assay data record.

5   As described herein below, the methods herein can be used for processing decoding and assay images, for extracting representations suitable for rapid transmission and for rapidly and reliably combining decoding and assay image signatures so as to associate assay results recorded from specific array locations with corresponding chemically encoded probe identities.

10   **Assembly of Random Encoded Bead Arrays.** Random encoded arrays may be assembled by the methods described in US Patent No. 6,251,691, or in US Patent Application Serial No. 10/192,352, entitled "Arrays of Microparticles and Methods of Preparation Thereof," incorporated herein by reference in its entirety. These methods combine separate batch processes that respectively serve to produce

15   application-specific substrates (e.g., chips at the wafer scale) and encoded bead libraries whose constituent beads are functionalized (e.g., at the scale of $\sim 10^8$ beads/100 ul of suspension) to display receptors, such as nucleic acids and proteins of interest. Beads assembled in an array may be immobilized by physical or chemical means to produce fixed random encoded arrays.

20   In addition, the methods described in US Patent No. 6,251,691 may be used to form multiple bead arrays. Alternatively, multiple bead arrays can be formed simultaneously in discrete fluid compartments maintained on the same chip. The integration of array assembly with microfluidics produces a self-contained,

14

miniaturized, optically programmable platform for parallel protein and DNA analysis. Once formed, these multiple bead arrays may be used for concurrent processing of multiple samples.

Spatial encoding of multiple arrays also can be accomplished by assembling planar bead arrays in a desired location, using discrete fluid compartments or the assembly methods described in US Patent No. 6,251,691. Alternatively, spatial encoding can be accomplished by assembling separate chips, each carrying at least one random encoded array drawn from a specific pool, into a designated configuration of multiple chips.

Chemical Encoding and Functionalization of Beads. Chemical encoding may be accomplished by staining beads with sets of optically distinguishable tags, such as those containing one or more fluorophore dyes spectrally distinguishable by excitation wavelength, emission wavelength, excited-state lifetime or emission intensity. Two-color and three-color combinations, where the latter may be constructed as "stacked" two-color combinations, are decoded as described herein.

The optically distinguishable tags may be used to stain beads in specified ratios, as disclosed, for example, in Fulwyler, US Patent No. 4,717,655, which is incorporated herein by reference in its entirety. Staining may also be accomplished by swelling of particles in accordance with methods known to those skilled in the art (*see, e.g.,* Molday, Dreyer, Rembaum & Yen, J. Mol Biol 64, 75-88 (1975); L. Bangs, "Uniform latex Particles, Seragen Diagnostics, 1984). Beads can be encoded by swelling and bulk staining with two or more colors, each individually at separate intensity levels, and mixed in separate nominal molar ratios in accordance with

15

methods known to the art. *See also* U.S. Patent Application Serial No. 10/348,165, entitled: Method of Controlling Solute Loading of Polymer Microparticles; filed 1/21/2003. Combinatorial color codes for exterior and interior surfaces is disclosed PCT/US98/10719, which is incorporated herein by reference.

5   Beads to be used in the bead arrays of the invention for biomolecular analysis are functionalized by a binding agent molecule attached thereto, where the molecule may be, for example, DNA (oligonucleotides) or RNA, fragments, peptides or proteins, aptamers and small organic molecules attached in accordance with processes known in the art, *e.g.,* with one of several coupling reactions of the known

10   art (G. T. Hermanson, *Bioconjugate Techniques* (Academic Press, 1996); L. Illum, P. D. E. Jones, *Methods in Enzymology* 112, 67-84 (1985)). The binding agent molecule may be covalently attached to the bead. Beads may be stored in a buffered bulk suspension until needed.

        Functionalization typically may be performed, for example, with a one-step

15   or two-step reaction which may be performed in parallel using standard liquid handling robotics and a 96-well format to covalently attach any of a number of desirable functionalities to designated beads. Beads of core-shell architecture may be used (as described in U.S. Provisional Application entitled: "Ionic Gel-Shell Beads with Adsorbed or Bound Biomolecules," filed 10/28/2003, and applications

20   claiming priority thereto) where the shell is a polymeric layer. Samples may be drawn along the way for automated QC measurements. Each batch of beads preferably has enough members such that chip-to-chip variations with different beads on chips are minimized.

16

Beads may be subjected to quality control (QC) steps prior to array assembly, for example, the determination of morphological and electrical characteristics, the latter including surface ("zeta") potential and surface conductivity. In addition, assays may be performed on beads in suspension before they are introduced to the substrate, to optimize assay conditions, for example, to maximize assay sensitivity and specificity and to minimize bead-to-bead variations. QC steps for substrates may include optical inspection, ellipsometry and electrical transport measurements.

Substrates. Substrates, *e.g.*, silicon wafers and chips, are used which may be patterned by invoking standard methods of semiconductor processing, for example to implement interfacial patterning methods of LEAPS by, *e.g.*, patterned growth of oxide or other dielectric materials to create a desired configuration of impedance gradients in the presence of an applied AC electric field. See U.S. Patent No. 6,251,691. Patterns may be designed so as to produce a desired configuration of AC field-induced fluid flow and corresponding particle transport, or to trap particles in wells, as described in US Provisional Application entitled: Immobilization of Bead-displayed Ligands on Substrate Surfaces," filed 6/12/2003, Serial No. 60/478,011.

In addition, substrates may be compartmentalized by depositing a thin film of a UV-patternable, optically transparent polymer to affix to the substrate a desired layout of fluidic conduits and compartments to confine fluid in one or several discrete compartments, thereby accommodating multiple samples on a given substrate. Other substrates such as patternable or machinable ceramics also are suitable.

Customization by Pooling. Bead-displayed probes of interest can be selected from a library of beads and pooled prior to array assembly. That is, customization of assay composition is achieved by selecting aliquots of designated encoded beads from individual reservoirs in accordance with a specified array composition. Aliquots of pooled suspension are dispensed onto a selected substrate (e.g., a chip). The aliquots may be mixed or may be separated to form a multiplicity of planar random subarrays of encoded beads, each subarray representing beads drawn from a distinct pool. The array may be laid out in a manner such that aliquot positions in the array correspond to the identity of each aliquot of the pooled bead population.

Array Analysis. The binding interaction between the receptor (which may be an oligonucleotide) displayed on color-encoded functionalized beads and a ligand (or "analyte") may be analyzed after a random encoded bead array is assembled in a designated location on the substrate or chip. For example, bead arrays may be formed after completion of the assay, subsequent to which an assay image and a decoding image may be taken of the array.

Immobilization Microparticle arrays may be immobilized by mechanical, physical or chemical anchoring as described in PCT/US01/20179 (counterpart of U.S. Patent Application Serial No. 10/192,352), including by trapping particles in wells, as described in US Provisional Application entitled: Immobilization of Bead-displayed Ligands on Substrate Surfaces," filed 6/12/2003, Serial No. 60/478,011.

In certain embodiments, bead arrays may be immobilized by physical adsorption mediated by application of a DC voltage, set to typically < 5V (for beads in the range of 2-6 Ym, a gap size of 100-150 Ym, and a silicon oxide layer of ~100

18

Angstrom thickness). Application of such a DC voltage for < 30s in "reverse bias" configuration - so that an n-doped silicon substrate would form the anode - causes bead arrays to be permanently immobilized. *See* US Patent No. 6,251,691.

In certain embodiments, the particle arrays may be immobilized by chemical means, *e.g*, by forming a composite gel-particle film. In one exemplary method for forming such gel-composite particle films, a suspension of microparticles is provided which also contain all ingredients for subsequent in-situ gel formation, namely monomer, crosslinker and initiator. The particles may be assembled into a planar assembly on a substrate by application of the LEAPS^{TM} process, as described in US Patent No. 6,251,691. Following array assembly, and in the presence of the applied AC voltage, polymerization of the fluid phase is triggered by thermally heating the cell ~ 40-45^{0}C using an IR lamp or photometrically using a mercury lamp source, to effectively entrap the particle array within a gel. Gels may be composed of a mixture of acrylamide and bisacrylamide of varying monomer concentrations, from about 20% to 5% (acrylamide : bisacrylamide = 37.5 : 1, molar ratio), or, in the alternative, any other low viscosity water soluble monomer or monomer mixture may be used as well. In one example, thermal hydrogels are formed using azodiisobutyramidine dihydrochloride as a thermal initiator at a low concentration ensuring that the overall ionic strength of the polymerization mixture falls in the range of ~ 0.1mM to 1.0mM. The initiator used for the UV polymerization is Irgacure 2959$^®$ (2-Hydroxy-4'-hydroxyethoxy-2-methylpropiophenone, Ciba Geigy, Tarrytown, NY). The initiator is added to the monomer to give a 1.5 % by weight

solution. The methods described in U.S. Patent Application Serial No. 10/034,727 are incorporated herein by reference.

In certain embodiments, the particle arrays may be immobilized by mechanical means, for example, such arrays may be placed into an array of recesses

5 may be produced by standard semiconductor processing methods in the low impedance regions of the silicon substrate. The particle arrays may moved into the recesses by, *e.g.*, utilizing LEAPS™-mediated hydrodynamic and ponderomotive forces to transport and accumulate particles in proximity to the recesses. The A.C. field is then switched off and particles are trapped into the recesses and mechanically

10 confined. Excess beads are removed leaving behind a geometrically ordered random bead array on the substrate surface.

Carriers and Cartridges. Substrates (*e.g.*, chips) with immobilized bead arrays may be placed in distinct enclosed compartments, and samples and reagents may be transported in and out of the compartments by means of fluidic

15 interconnection. On-chip immunoassays, including those for various cytokines, *e.g.*, interleukin (IL-6) may be performed in this format. In such immunoassays, samples are allowed to react with beads immobilized on the chip and adsorption of targets in the samples by the receptors on the beads may be detected by binding of fluorescently labeled secondary antibodies.

20 **Random Encoded Array Detection.** Once the functionalized and encoded beads are prepared and assembled on the substrate, a binding assay may be performed. The array can function as a two-dimensional affinity matrix which displays receptors or binding agents (*e.g.*, oligonucleotides, cDNA, aptamers,

20

antibodies or other proteins) to capture analytes or ligands (oligonucleotides, antibodies, proteins or other cognate ligands) from a solution or suspension that is brought in contact with the array. The bead array platform may be used to perform multiplexed molecular analysis, such as, *e.g.*, genotyping, gene expression profiling, profiling of circulation protein levels and multiplexed kinetic studies, and may be used for the implementation of random encoded array detection (READ™), including analysis based on image acquisition, processing and analysis.

I.2 Multicolor Image Acquisition Using an Automated Array Imaging System

Multicolor images can be used to encode and display information recorded in two or more color channels. The construction of multicolor images can be accomplished by merging two or more images recorded in separate spectral bands and distinguished by selection of suitable color filter combinations, as is well-known in the art.

The READ™ format provides for multicolor images before and after the assay, referred to respectively as a decoding image and an assay image. The decoding image serves to record the location of particular identified solid phase carriers - and hence the identity of receptors displayed on such carriers – based on their color-encoding in an array. These solid phase carriers may be color encoded using, for example, combinations of two or more fluorescent dyes. The assay image reflects the optical signatures induced by association of target analytes with carrier-displayed receptors. In one example, useful in gene expression profiling, signal produced on an array by the hybridization of cDNA or RNA produced from a tissue

21

sample of interest and labeled with fluorescent dye (e.g., Cy5) may be compared

with the signal produced from a known quantity or concentration of cDNA or RNA

produced from a reference sample and labeled with a fluorescent dye, (e.g., Cy3 or

Cy5). The comparison of signal from the tissue sample and the reference sample

indicates the level of gene expression. An analogous format may be used in

molecular cytogenetics applications. Protein-protein interactions can also be

monitored with this format, where the protein in the sample is labeled following its

association with the bead-bound protein in the array, for example, by using a labeled

antibody which targets the protein.

Multicolor images obtained from monitoring of random encoded arrays can

be recorded automatically. The arrays can be formed on chips, and multiple chips

can be placed on a carrier, such as that shown in Fig. 1. However, alternative modes

of presenting and arranging random encoded arrays are possible. For example,

samples also may be mounted in flow cells or cartridges permitting fluidic operation

so as to inject samples, reagents and buffers, permitting imaging of probe array by

way of standard microscope optics.

The components and subsystems of an exemplary image acquisition system

may include the following:

Input/Output File System: images are handled in TIFF format, other files are handled

in XML (eXtensible Markup Language) format; an XML output file records the

settings of parameter such as image acquisition integration time, filter selection

System Status: illumination source (ON/OFF), stage target position

Mechanical Subsystems: xy translator, z actuator, filter wheel, ND filters

22

Barcode Reader

Image Acquisition and Storage

Control Software

Graphical User Interface (GUI)

5   Autocenter Function (implemented by Software)

Autofocus Function (implemented by Software)

Hardware. Fig. 4 shows a microscope, for use in an imaging system, providing transmission or reflection geometry and multiple methods of generating optical contrast. In one embodiment, reflection geometry is chosen to record "brightfield"

10 (*e.g.,* an image not recorded under fluorescence optical contrast imaging conditions, including an image recorded under white light illumination) as well as multicolor fluorescence images in a fully automated manner.

Other components and subsystems of an imaging system are set forth below.

Illumination Source. Depending on the application of interest, any suitable

15 illumination source (400) can be used, including a laser, or a standard microscope illumination sources including tungsten halogen, mercury and xenon. In one embodiment, a xenon light source is used for multi-fluorescence imaging. A mechanical shutter (410) controls "Light-ON" and "Light-OFF" functions.

Mechanical Subsystems. Certain subsystems can be used to control precision

20 positioning of the sample (420), selection of image mode, namely brightfield or (epi)fluorescence (using different filters), and spectral filter selection. The positioning of the sample involves horizontal and vertical sample positioning, deploying a computer-controlled xy-translator (part of the xyz stage (430)), which

23

can be under the control of a manual "joystick" positioning function or an automated autocentering function, and a z-actuator connected to a vertical motion of the sample under control of an autofocusing function which may be computer-controlled. Fluorescence filter combinations can be selected automatically using a computer controlled carousel housing filter cubes (440).

Barcode Scanner. A handheld barcode scanner (450) can be used to scan an identifyng barcode affixed to each multichip carrier or other sample carrier or cartridge. The barcode can identify, for example, the composition of beads associated with the carrier, the origin of the carrier (i.e., the batch it is derived from) or other information.

Optical Subsystem – A combination of microscope objective (460) and collection optics (470) in standard configuration, or in Koehler configuration, is used for illumination as well as collection and image formation.

CCD Camera. A CCD camera (480), preferably with C-mount, is attached to the microscope to record images.

Control Software. The fully automated operation of the array imaging system is enabled by control software (also referred to herein as the Array Imaging System - Operation Software (*"AIS-OS"*)) comprising a Graphical User Interface (GUI) as well as control algorithms implementing autocentering and autofocus functions, as set forth below.

Operation of Array Imaging System. Bead arrays mounted in a multichip carrier or cartridge are placed on the translation stage of the Array Imaging System ("AIS") and multicolor images (both decoding and assay images) are recorded. Table I below

24

shows the pseudocode for the translation/decoding operation of the Array Imaging

System.

Table I

```
IF (record Decoding Image mode) {
        LoadFile (Production Data Record);  /** Load or Create **/
        ScanBarCode (Carrier ID);
        WriteInfo (Carrier ID, Number of BeadChips on Carrier,
        ProductionDataRecord);
        };


ELSE IF (record Assay Image mode) {
        ScanBarCode(Carrier ID);
        ReadInfo (Production Data Record, CarrierID, Number ofBeadChips on
        Carrier);
                        / ** The Production Data Record is accessible as an
                        XML file in the folder set   up by the AIS for each
                        sample **/


        ConstructFileName (AssayDataRecord)      /** to match Production Data
        Record **/
        WriteInfo(Assay Data Record, CarrierID, Number of BeadChips on Carrier);
                        / ** Production Data Record and Assay Data Record
                        are maintained on an SQL database server accessible
                        to authenticated users **/
        };


        index = 0;
        WHILE( index < Number of BeadChips on Carrier)
        {
                IF (record Assay Image mode) {
                        ReadTargetPosition(Production Data Record, BeadChipID, X, Y, Z);
                };


                MoveXYTranslator ( X, Y );MoveZTranslator (Z);
                Z = AutoFocus(Z);                       /** perform autofous operation **/
                AutoCenter(X, Y);              /** In record Decoding Image mode,
                        operator sets first center position **/
                Z = AutoFocus(Z);                       /** repeat autofocus **/


                IF (record Decoding Image mode) {
                        WriteTargetPosition(Production Data Record, BeadChipID, X, Y, Z);
```

25

```
          };

          FOR (each desired ColorChannel) {
                  SetFilter (ColorChannel);
   5              AcquireImage ( );                  /** See details below **/
                  WriteImageFile(CarrierID, BeadChipID, ColorChannel, ImgFile);

                  IF (record Decoding Image mode) {
                          UpdateProductionDataRecord(CarrierID, BeadchipID,
  10                      ImgFileName);
                  };
                  ELSE IF (record Assay Image mode) {
                          UpdateAssayDataRecord(CarrierID, BeadChipID,
                          ImgFileName);
  15              };
          };
          index++ ;
   };
```

20   **Autocentering.** The autocentering function, using a given input image, positions the

      XY translation stage so as to place each selected array into the center of the imaging

      system's field view by determining that the image in the viewing field is, in fact, a

      rectangle with the correct number of sides and right angle corners.   This is

      accomplished by performing the steps in Table II, which shows the pseudocode for

25   the autocentering operation:

                                      Table II

_____

```
30    OpenImg(InputImg, OpenedImg);    /** apply sequence of morphological erosion
      and  dilation operations to eliminate internal structure of the image showing particle
      array **/

      Binarize(OpenedImg,BinImg);                      /** apply optimal thresholding
35    algorithm **/.

      CloseImg (BinImg, ClosedImg);                    /** apply sequence of
      morphological  dilation and erosion operations **/
```
                                         26

```
AnalyzeConnectivity (ClosedImg, ConComp);          /** find connected
components in closed image **/


Filter (ConComp, FilteredConComp);                  /** filter out all "non-
box-like" regions; a "box-like" region is defined as a region whose area is close to
the area of its "bounding box" **/


Center = FindMaxConComp(FilteredConComp);   /** find largest connected
component that is smaller than 70% of the image size and find its centroid **/


MoveXYTranslator ( Center.X, Center.Y );          /** position stage **/
```

Only when a new multichip carrier ("MCC") is first inspected and decoded, does

the positioning of the very first array on the MCC require interactive operation.

This initial positioning step is performed as a part of array assembly or subsequent

quality control. All subsequent positioning may be automatic. The full processing-

positioning followed by acquiring of multiple images and displaying a rendering of

a multicolor image typically requires only a few minutes.

Autofocusing. The autofocusing function positions the Z actuator so as to bring the

image in focus and place each selected array into the center of the imaging system's

field view. An algorithm which uses a local contrast function to determine optimal

focus can be used. This local contrast is determined as follows using a fast

computation: evaluate, for each pixel, a quantity Δ_{max}, defined as the largest

absolute value of the difference in intensity between that pixel and its four

horizontal and vertical neighbors; next, sum the Δ_{max} over a designated portion of

the image: this serves to speed up the operation. The Z-position of maximal contrast

27

is located. The autofocus function should help ensure vertical positioning to within one micron or less.

I.3 Performing Multi-Analyte Molecular Analysis

The process of performing multianalyte molecular analysis using the system herein would, in an exemplary embodiment, involve the concatenation of the previously described operations as follows (Fig. 5):

Collect patient sample (510)

Transfer to sample container, preferably a barcoded sample container (520)

Process sample (530) using requisite reagents (540) to produce analyte (550)

Select multichip carrier (MCC), obtain MCC information (560)

(580), for example, positions of beadchips arranged on MCC;

Perform assay (590) to produce transformed analyte (591) to be analyzed

Mount MCC in array imaging system and read MCC barcode (570) to obtain assay configuration

Acquire assay image(s) (592)

Submit assay image data for processing and analysis, details of which are described below.

II Image Processing and Analysis

Processing, analyzing, transmitting and storing images as set forth herein can be implemented in Visual C++ (MicroSoft) using a graphical user interface software package including .exe files implementing the pseudocodes and flow diagram steps

28

described herein including, for example, the analysis, processing and decoding steps described herein.

An image processing program (designated "DECODER"), which can be run, for example, on the Microsoft Operating system, e.g., Windows 98 or 2000, and

5   which contains functions to display, process, save and print "multicolor" sets of multiple microarray images in an integrated graphical user interface (GUI) and to generate a decoding data record which may be submitted for further analysis to the ANALYZER, residing either on the same computer or on a separate computer. This program can be readily implemented by those skilled in the art, using the outlines

10   herein. As illustrated in Fig. 6, the operation of the DECODER comprises reading the Decoding Image Record (610, 620), rendering and displaying decoding image(s) (630), and processing decoding image(s) (640) to create a Decoding Data Record (650), display scatter plots (660) and update databases (670).

Another image processing program (designated READER), which can be run,

15   for example, on the Microsoft Operating system, e.g., Windows 98 or 2000, and functions to display and to process pairs of assay images acquired so as to generate an assay data record which may be submitted for further analysis to the ANALYZER, residing either on the same computer on a separate computer. As illustrated in Fig. 7, the operation of the READER comprises reading the Assay

20   Image Record (710, 720), rendering and displaying assay image(s) (730), and processing assay image(s) (740) to create an Assay Data Record (750) and update databases (760).

ANALYZER is an analysis program. As illustrated in Fig. 8, ANALYZER receives input from DECODER, in the form of a Decoding Data Record (810), and READER, in the form of an Assay Data Record (820), and comprises functions to perform further analysis including: cluster analysis (830) to create a decoding map (840). The decoding map is combined (850) with an assay data record so as to produce a Decoded Assay Data Record (860) which is displayed (870) in a variety of formats and stored in a database (880).

ANALYZER, DECODER and READER may reside on separate computers which may communicate by way of a data network. In this manner, data from assays can be received from a remote site but can be decoded and analyzed at another site. In such embodiment, DECODER and ANALYZER may be integrated into a single program or loaded onto the same computer.

IChipReader provides a COM interface in the form of a dynamically linked library linking the functions of DECODER and READER.

This image analysis system has the following advantages.

Reliability: Robust algorithms have been designed in order to handle images of widely varying quality encountered in practice, including images exhibiting very low contrast or variations in contrast across the image, significant noise and corruption of edges or features, or displacement and misalignments between multiple images of a given array. These robust algorithms ensure the reliability of the results produced by the analysis.

Accuracy: The entire sequence of processing steps is performed without human intervention, thereby avoiding error and enhancing ease-of-use.

Speed: Algorithms have been designed for efficiency, and functions have been integrated so as to minimize processing time. A chip displaying a single array can be processed in as little as 4 seconds.

Productivity and throughput: Sets of images may be analyzed in batch mode.

Ease-of-use and convenience: The GUI package provides convenience and flexibility in controlling all system functions.

Functions and capabilities provided by these systems including processing, analyzing, transmitting and storing images are elaborated below.

II.1   Array Segmentation and Extraction of Signal Intensities

Image processing may be applied to each of the one or more constituent images of a composite image in the decoding data record or assay data record to segment the image and extract a textual representation of the signal intensity distribution within the array. This representation would serve as input for further analysis.

In certain embodiments, decoding and assay images or the corresponding data records are analyzed to obtain quantitative data for each bead within an array. The analysis invokes methods and software implementing such methods to: automatically locate bead arrays, and beads within arrays, using a bright-field image of the array as a template; group beads according to type; assign quantitative intensities to individual beads; reject processing "blemishes" such as those produced by "matrix" materials of irregular shape in serum samples; analyze background intensity statistics; and evaluate the background-corrected mean intensities for all bead types along with the corresponding variances.

31

II.1.1 Referenced Arrays

Referenced arrays are located in designated positions, and in designated orientations, with respect to features designed into patterned substrates in accordance with methods previously disclosed in PCT/US01/20179, U.S. Patent No. 6,251,691 and U.S. Patent Serial No. 10/192,352. For example, a locus of low impedance on a substrate may be designed to collect particles using the LEAPS™ method and may further contain a central recess grid to mechanically immobilize microparticles. *See* U.S. Patent No. 6,251,691.

Specifically, in one embodiment, following completion of AutoCentering and AutoFocusing as described above, the system makes a record, in both the "record Decoding Image" and the "record Assay Image" modes, of both the brightfield image and one or more color images of the array, the color images being recorded following selection of the desired filter settings as described above. In one embodiment, color images in the "record Decoding Image" mode are recorded in a "BLUE" and in a "GREEN" channel selected by respective filter combinations (Blue Channel: excitation filter : 405nm (20nm); emission filter : 460nm (50nm) and beam splitter : 425nm (long pass), Green Channel: excitation filter : 480nm (20nm); emission filter : 510nm (20nm) and beam splitter : 495 nm (long pass)) and one color image in the "record Assay Image" mode is recorded in a "RED" channel selected by a filter combination Red Channel (excitation filter : 640nm (30nm); emission filter : 700nm (75nm) and beam splitter : 660nm (long pass)). The excitation filters transmit only those wavelengths of the illumination light that efficiently excite a specific dye, and an emission filter attenuates all the light transmitted by the

32

excitation filter and transmits any fluorescence emitted by the specimen, and a beam splitter reflects the excitation light but transmits the emitted fluorescence (the figures in parenthesis indicate the width of the band for each filter). Therefore, the AIS system, in either "record Decoding Image" mode or "record Assay Image" mode,

5   permits recording of images in two or more color channels.

Processing of images recorded from referenced arrays may be performed by extracting a reference "mesh" or "grid" structure, where individual fields in the grid include beads and the outer dimensions of the grid correspond with the dimensions of the referenced array. The principal operations common to the processing of

10   Decoding Images and Assay Images include segmentation to locate array boundaries, mesh/grid delineation, image registration (or alignment) and extraction of intensities, as elaborated below. Following completion of processing steps, further analysis is performed by constructing a decoding map from two or more decoding images using a cluster algorithm and by merging decoding and assay data record to produce

15   decoded assay data record. Partial or complete results and related information may be exchanged between two or more parties, as further elaborated herein.

In the *"process Decoding Image"* mode, the following operations, as shown in Table III, are utilized. Table III shows the pseudocode for an exemplary processing of a decoding image.

20
Table III

LoadImage (BrightField Image);
FindBoundary (BrightFieldImage, RotAngle);   /** using brightfield
25   image, find array boundary and angle of misorientation with respect to display edges
**/

33

```
          RotateImage(BrightFieldImage, RotAngle);
          FindGrid (BrightFieldImage, Grid)        /** locus of local intensity minima **/

5         FOR (each decoding image be processed) {
                  LoadImage (Fluorescence Image);
                  RotateImage (Fluorescence Image,  RotAngle);
                  AlignImage(FluorescenceImage);    /** align fluorescence image with bright
                  field image **/
10                OverlayGrid ( );

                  ReadIntensityDistrib(DecodingDataRecord, SampleMesh, Grid,
                  FluorescenceImage);
          }
```

These steps are followed by the step of creating a scatter plot from two or more

decoding images and performing a cluster analysis to establish a decoding map.

These steps are elaborated below.

In the *"process Assay Image"* mode, the pseudocode for the processing of an assay

image, as illustrated in Table IV, are used.

Table IV

```
          LoadImage (BrightField Image);
          FindBoundary (BrightFieldImage, RotAngle );    /** using brightfield image, find
          array boundary and angle of misorientation with respect to display edges **/

25
          RotateImage(BrightFieldImage, RotAngle);
          FindGrid (BrightFieldImage, Grid)              /** locus of local intensity
          minima **/

30        FOR (each assay image to be processed) {
                  LoadImage (AssayImage);
                  RotateImage (Assay Image,  RotAngle);
                  AlignImage(AssayImage);                /** align assay image with bright field
                  image **/
35                OverlayGrid ( );

                  ReadIntensityDistrib (AssayDataRecord, SampleMesh, Grid, AssayImage);

          }
```

Histogram expansion is applied to brightfield and color images prior to finding boundaries, locating grids and aligning images; only the intensity extraction is performed on the 16-bit image as originally recorded. The principal operations are implemented using standard methods (Seul, O'Gorman & Sammon, "Practical Algorithms for Image Analysis", Cambridge University Press) as follows.

FindBoundary - Array edges in the brightfield image (and optionally any of the color images) are located using a standard Sobel y-gradient operator image for left and right edges and a standard Sobel x-gradient operator for top and bottom edges. Using these edges, the location of the array and its misorientation with respect to the image display boundaries are computed. For future use, the array is rotated to bring it into alignment by with the image display. Prior to edge detection, noise is filtered by applying six iterations of a morphological "Open" operation, which is an image processing technique. See Seul, O'Gorman& Sammon, *supra*.

Mesh/Grid Construction - A grid or mesh delineating intensity *maxima* ("peak") is extracted by tracing the locus of local intensity *minima* ("valley") within the brightfield image. This locus defines a mesh or grid such that each field in the mesh delineates a local intensity maximum associated with a bead or with a recess provided in patterned substrates. In this manner, the grid traces around each of the beads, and includes one bead in each segment of the grid.

To implement the mesh construction, the problem is mapped to Dijkstra's "shortest path" algorithm (*see* Introduction to Algorithms, T. Cormen, C. Leiserson et al., The MIT Press) well known in the fields of computational geometry and combinatorial optimization, by ascertaining the intensities of image pixels with

35

values of vertices in a graph. The algorithm finds the mesh, also referred to as a grid, as an optimal path as follows:

Pre-process Image:

Compute the external gradient image by subtracting a dilated image from the original image.

Determine overall orientation, either horizontal or vertical

As illustrated in Fig. 15A, provide horizontal and vertical reference lines of known length to determine the orientation of the underlying hexagonal grid, compute shortest paths for the two reference lines and compare to corresponding reference line lengths by forming. The ratio closes to unity indicates either horizontal or vertical orientation.

Find horizontal grid partition:

Replicate reference lines by shifting by unit mesh size, then compute shortest path to determine actual grid line. Continue until replicated line falls outside array boundary.

Find vertical grid partition:

As illustrated in Fig15b, find the shortest path of a diagonal line - depending on overall orientation - oriented at either 30 degrees or 60 degrees with respect to the horizontal lines in the case of an anticipated underlying hexagonal lattice of average intensity maxima. Compute intersections of the diagonal lines and every horizontal line given the intersections of the diagonal lines and two consecutive horizontal lines, vertical lines will be located at the midpoint of these intersections.

36

Post-process grid:

Grow or shrink grid defined by the totality of the horizontal and vertical lines to the expected array boundary; correct grid stagger.

Store grid:

Store grid coordinates in a file.

Image Registration - Given a grid, a registry of the brightfield image with one or more color images ensures proper alignment by eliminating possible misorientation and translation ("shifts") between the multiple images recorded from a given array. One source of such shifts is the wavelength-dependent refraction introduced by standard fluorescence filter combinations. Registration aligns the assay and decoding images.

Misorientation is eliminated by rotation to bring a given image into alignment with a reference such as the brightfield image grid. An alternative for aligning images without reference to a brightfield image is described below.

Assuming only translational displacement, the system can invoke the following fast algorithm. To determine horizontal displacement ("X-shift"), construct intensity profiles along vertical scan lines within the array boundary; similarly, to determine vertical displacement ("Y-shift"), construct intensity profiles along horizontal scan lines within the array boundary. Next, construct horizontal and vertical profiles along lines displaced from the first set by one respectively one horizontal or one vertical mesh unit. The peak in the profile determines the image shift.

The methods herein are limited to shifts between images of less than half of the mesh size, by the optical subsystems of the system including the CCD camera used herein. If larger shifts are encountered, image registration may be off by one mesh unit in row and/or column dimensions. A Minimal Variance Matching

5   algorithm described below is utilized to correct larger misalignments.

Intensity Extraction. Following completion of image registration, in one embodiment, intensities are extracted from each color image by sampling the interior, and not the exterior, of each field of the mesh/grid with an averaging filter mask of suitably chosen size to fit into the interior of each mesh field. Each intensity

10   value is optionally corrected by subtracting a supplied background value. One method of supplying the background values is to record it as an average of pixels values from an area of the image outside the mesh/grid boundary.

In contrast to widely used conventional methods that invoke peak finding and peak fitting algorithms to locate object positions, the present method offers

15   substantial advantages of processing speed.

Extracted intensities are stored - optionally in binary form - in a one-dimensional array of length L, L denoting the number of units of the grid/mesh constructed in the course of segmentation. This one-dimensional array can be mapped onto the grid or mesh to associate each intensity value with a unique

20   coordinate within the bead array. For example, the following structures may be used:

```
float intensity[4012];        /** array holding 4012 intensity values **/

Struct Grid{

            int leftUpX,
```

int leftUpY,

int rightDownX,

int rightDownY}; /** structure representing one grid field **/

Grid grid[4012];   /** array holding 4012 grid fields **/

5

Eliminating Reference to Brightfield Image. In certain embodiments, one may eliminate reference to the brightfield image in the course of image processing, notably during the step of eliminating image misalignment, and indeed to eliminate the step of recording the brightfield image altogether. In that case, the step of

10   aligning color images with the display boundaries invokes information extracted directly from the color images.

The approach is conceptually as follows. Given a color image, construct horizontal and vertical intensity profiles by respectively projecting image intensities to the top-most and left-most scan line in the display, then evaluate the intensity

15   variation in each profile. Next, rotate the color image by a pre-defined angle and repeat the previous construction. Continue to rotate until the profiles exhibit maximal variations, then reduce the step size in rotation angle and reverse the direction of rotation until the optimal rotation angle is found.

This procedure is significantly improved when information about the array

20   geometry is available a priori. For example, in one embodiment of referenced arrays, a hexagonal geometry with specific choice of nearest neighbor separation, a, and alignment of principal axes with the chip edge is chosen. Then, the desired alignment is characterized by one of the profiles assuming the form of a periodic variation with

39

a single periodicity, a, and the other profile assuming the form of a superposition of two phase-shifted periodic functions, both with periodicity, a*cos 30°. Horizontal and vertical profiles produced by such an array at a given misalignment angle thus may be analyzed by fitting each to a superposition of two trial functions and obtaining the angle of misalignment from the fit.

II.1.2 "Non-referenced" Arrays

Decoding and Assay Images. To perform a multiplexed binding assay in accordance with the READ™ process, the array is first imaged by multicolor fluorescence, to determine the color code of constituent beads which uniquely correspond to the chemical identity of the probe displayed on the bead surface; second, to record the fluorescence intensity which indicates the amount of probe-target complex formed on each bead surface in the course of the binding or hybridization assay. The process of image detection and bead decoding is described in PCT/US01/20179 (WO 01/98765), incorporated herein by reference in its entirety.

Quantitative Analysis. Image analysis algorithms that are useful in analyzing the data obtained with the READ process disclosed herein may be used to obtain quantitative data for each bead within an array, as set forth in PCT/US01/20179, incorporated herein by reference. In preferred embodiments, data are obtained from the decoding and the assay images, or preferably from the corresponding decoding image record and assay data record by application of certain algorithms. These algorithm may be used to obtain quantitative data for each bead within an array. The analysis software automatically locates bead centers using a bright-field image of the array as a template, groups beads according to type, assigns quantitative intensities to

40

individual beads, rejects "blemishes" such as those produced by "matrix" materials of irregular shape in serum samples, analyzes background intensity statistics and evaluates the background-corrected mean intensities for all bead types along with the corresponding variances. Using calibration beads that are included in the assay, intensities are converted to an equivalent number of bead-bound fluorophores.

II.2   Representation and Encryption of Array Configurations - ChipID

II.2.1 Covering. Given a set of probe molecules of types $P = \{p(1), ..., p(k), ..., p(n)\}$ and a set of tags, $T = \{t(1), ..., t(k), ..., t(n)\}$, the former, for example in the form of oligonucleotides of defined length and sequence, the latter for example in the form of color codes associated with a set of beads, one defines a one-to-one mapping of T onto P whose image represents a covering, $C := C(P)$ of the set P. The covering is obtained by attaching probes in set P to color-encoded beads in set T.

In certain embodiments, encrypted coverings serve to conceal the identity of probe molecules associated with tags by revealing, for each probe molecule, only a label or pointer that is logically linked to that probe molecule, but not the probe identity itself. This is disclosed only by a "de-covering" process.

II.2.2   Encoding Random Array Configurations

The random assembly of pooled beads of different types into a planar array creates a specific configuration, thereby defining a "random encoding", E, as follows. Given a set of tags, $T = \{t(1), ..., t(k), ..., t(n)\}$, for example in the form of color codes associated with a set of pooled particles, define E as the mapping of T onto a set of positions, $V = \{v(1), ..., v(l), ..., v(L)\}$, constructed as follows: from each of n reservoirs of particles, each reservoir containing particles that are uniquely

41

associated with one tag in accordance with $T = \{t(1), ..., t(k), ..., t(n)\}$, draw $r(k)$ (indistinguishable) particles and place them into $r(k)$ positions randomly selected from a set $V = \{v(1), ..., v(l), ..., v(L)\}$. In a preferred embodiment, V corresponds to the vertices of a rectangular array, $\{(i, j); i = 1, ..., I, j = 1, ..., J\}$ or, equivalently, $\{l; l = 1, ..., L: = I*J\}$, of designated positions ("traps") in a silicon substrate.

In certain embodiments, encoding serves as a further level of encryption to conceal the identity of tags which is revealed only by the decoding process. In addition, standard encryption techniques may be applied to further conceal encoding and covering information. Decoding of the array configuration identifies the tag assigned to each of the positions within V. For example, each such color code, identified by a unique tag index, may be obtained by combining fluorescent dyes of fundamental colors, R, G and B, for example, in specified ratios to produce beads producing fluorescence signals of intensities (I_R, I_G, I_B), for example, in the respective color channels, R, G and B. The system described herein can maintain this information in a separate configuration file, which is generated in conjunction with the production of bead libraries.

Representations. In one embodiment, encoding is achieved by creating spectrally distinguishable particles by way of staining them with two or more dyes in accordance with one of several possible possibilities. For example, several fluorophore tags in the form of dyes, Red (R), Green (G) and Blue (B), for example, may be combined in a variety of fixed R-G-B molar ratios or may be combined in binary (or other) fashion, each dye being either present or not present in any given particle type. Decoding of an array of color-encoded particles is performed as

42

described herein by recording a set of images showing fluorescence intensities in separate color channels for each of the fundamental dyes and determining molar ratios by analyzing intensities in the various color channels.

The information from multiple decoding channels may be represented in a merged decoding intensity array which forms part of the Decoding Data Record described herein, by listing, for each position $v(l)$, $1 \ldots l \ldots L$, a set of intensities (I_R, I_G, $I_B)_l$, for example, or listing relative abundances that are obtained by normalizing intensities by suitable internal standards. Optionally, to obtain a compact integer representation, intensities may be represented in binary form, $I = 2^p$, $0 \leq p \leq 16$ so that a set of exponents (p_R, p_G, $p_B)_l$ may be stored for each position.

Further analysis, performed as described herein in susbsequent sections, serves to construct a decoding map. This map is composed of clusters, each cluster representing one spectrally distinguishable particle type which in turn is defined by a triplet of fundamental tags, such as (I_R, I_G, I_B). Once the decoding is in hand, clusters may be given a simple index which now serves as a tag index. That is, the triplet is replaced by a simple tag index. Accordingly, the random encoded configuration generated by E may be represented in the form of the random sequence of $L = \sum r(k)$ tag indices assigned by the encoding E to positions ($v(1)$, ..., $v(l)$, ..., $v(L)$). In certain embodiments, it will be convenient and useful to sort this sequential representation by tag index so as to obtain a one dimensional array of n lists, the k-th such list containing the sequence of $r(k)$ array positions occupied by tag k. If the positions are identified by the corresponding vertex array index, this provides a particularly compact representation.

43

Alternatively, a representation in the form of a 1-d array of length L of tag indices also may be convenient. For example, the configuration of an array composed of 4,096 or 212 beads of 128 = 27 types, could be stored in 4k*2Bytes = 8kB of non-volatile memory which could be packaged with the carrier.

5   **Individual Array Configurations: IntrinsicChipID**. In a preferred embodiment of a random encoded array assembled on a silicon chip substrate, the array configuration, in any of the aforementioned representations, provides an identifying tag for the substrate. See US Patent Application Serial No. 10/365,993 "Encoded Random Arrays and Matrices." filed 2/13/03, incorporated by reference. Each such

10   InstrinsicChipID is drawn from the number, S, of distinguishable configurations of a random encoded array of I*J = L vertices, given by the number of ways in which n (unordered) samples of r(k) (indistinguishable) particles of type t(k), $1 \leq k \leq n$, may be distributed among L positions:

15                          $$S (L; n; r(k), 1 \leq k \leq n) := L! / r(1)! \, r(2)! \, ... \, r(k)! \, ... \, r(n)!$$

Illustrating the large number of possible combinations is the fact that an array of L=16 positions, composed of n=4 bead types, where each type is represented four times (r(1) = ... r(4) = 4), can display $S(16; 4; r(k) = 4; 1 \leq k \leq 4) = 16! / (4!)^4$, or

20   approximately 63 million configurations. The InstrinsicChipID may be cast in any of the representations discussed above.

Degree of Randomness: Autocorrelation Function of Tag Sequence. Indeed, the degree of randomness of a given bead array is readily ascertained by constructing the

autocorrelation of the tag sequence corresponding to the random configuration of encoded beads within the array as elaborated above. For example, a random sequence of length 22 and composed of three tags, R, G, B with relative abundance $7/22 = 1/3$, will produce an autocorrelation function, g, of this type with the following behavior near the origin:

Shift: -1   R G G B R G R R B R G R B R G B B R G G B G

                     R G G B R G R R B R G R B R G B B R G G B G   g = 3

Shift: 0   R G G B R G R R B R G R B R G B B R G G B G

                     R G G B R G R R B R G R B R G B B R G G B G   g = 22

Shift: +1   R G G B R G R R B R G R B R G B B R G G B G

                     R G G B R G R R B R G R B R G B B R G G B G   g = 3

Scoring each tag match in the autocorrelation as 1, each tag mismatch as 0, it is readily seen that the (normalized) autocorrelation function of the random tag sequence will exhibit a sharp peak and will drop - within a single unit shift - to the average value of $\sim(1/r)^2$, r denoting the average redundancy of each tag. This property of random encoded arrays will serve to construct a robust "matching by variance minimization" algorithm to combine decoding and assay data records, as described herein.

II.3   **Image Analysis**

Completion of the aforementioned image processing steps yields a compact representation of the intensity distributions in decoding and assay images which facilitates further analysis. This analysis includes the steps of generating a decoding map from the set of decoding images and combining ("merging") decoding and assay images to generate final assay results using a matching algorithm further elaborated below.

II.3.1 Construction of Decoding Map by Cluster Analysis

A decoding map assigns each bead located in the processing of a decoding image to a unique group in accordance with its unique tag. For example, color-encoded beads will be grouped by color and/or by intensity of each of two or more encoding colors, as assumed here for clarity in the exposition of the clustering algorithm. It will be apparent that other codes are possible here and will be used in analogous fashion. The system herein may include two or more clustering algorithms.

II.3.1.1   Matching to Map Template

This algorithm anticipates a decoding map template, constructed manually or otherwise provided, which provides seed locations, each anticipated group or cluster in the decoding map corresponding to one such seed. This is particularly advantageous in the situations commonly encountered in practice involving analysis of a decoding image recorded from bead arrays of the same batch or lot. That is, the number of anticipated clusters, and their respective approximate central locations, are known *a priori*. Assuming, for purposes of illustration, ratio encoding by two

encoding colors, the algorithm produces a partition of a given scatter plot of decoding intensities which is first converted into a two-dimensional histogram.

The map template matching algorithm first generates a two-dimensional histogram image of the input data, optionally providing smoothing to the histogram image to eliminate noise. Given the two-dimensional histogram, the decoding group is generated using a "watershed" algorithm, well known in the art in connection with image segmentation, which treats the intensity histogram as a topographical map showing local elevation as function of position. Starting at the lowest point, the "water level" is now gradually increased until "water" starts to spread over two previously separate compartments. A "dam" is constructed at the "overflow" position. The set of dams so constructed represents the set of segment boundaries.

To implement these steps of generating the decoding map, the map template matching algorithm uses three auxiliary objects: a priority queue, a stage (of processing) image and a label image. The priority queue maintains individual pixels in accordance with their intensity values as obtained from the two-dimensional histogram of the input scatter plot, keeping the pixel with the maximal value at the top. The stage image serves to track the stage of pixel assignment: any given pixel either is or is not assigned to a group or cluster. The label image serves to track the group identity of each assigned pixel.

The algorithm proceeds as follows. For each given seed, initialize the corresponding pixel in the label image by assigning it the seed label, add its eight nearest neighbors to the priority queue and mark each of these pixels "assigned" in the stage image. Next, pop the top pixel from the queue and inspect its eight nearest

47

neighbors, ignoring unassigned pixels and checking whether all "assigned"

neighbors have the same label. If so, mark the pixel with that label, otherwise, leave

the pixel unmarked. Finally, add to the queue all non-zero neighbors not currently in

the queue and pop a new element. Continue until queue is empty - at which point the

5   label image shows the group assignment for each pixel.

The resulting partition assigns each data point in the scatter plot to one and

only one of the groups ("clusters") identified by the set of given seed locations. Two

or more of such sscatter plots are processed if three (or more) colors are used for

encoding. The algorithm performs the operations as illustrated in Table V. Table V

10   shows the pseudocode for the operation of constructing a Decoding Map by way of

template matching.

Table V

15
```
CreateScatterPlot(Image, IntensityArray(FirstColor), IntensityArray(SecondColor));
                    /** 2d Plot of Intensities extracted from Decoding Images in
                    different Selected color channels **/


20   ConvertScatterPlotToTwoDimHistogram(Image);
     SmoothImage(Image, Kernel);        /** apply 5*5 Smoothing Filter **/

     SetMapTemplate(MapTemplate); /** provide map template **/


25   GetSeedLocations(MapTemplate, ClusterSeeds);
     GenerateDecodingMap(ClusterSeeds);
```

30

II.3.1.2   Fast Clustering Algorithm

The system described herein also includes a fast algorithm that invokes graph theory to construct a two-dimensional decoding map without the aid of a template. The algorithm converts the input scatter plot of intensities into a "distance graph,"

5   each data point in the scatter plot representing one node in the graph, and each such node being connected by one edge to its K nearest neighbors (by Euclidean distance). Each edge is assigned a weight that is proportional to its length, and each node is given a value computed from the weight of the largest edge connected to that node.

10   The algorithm comprises the following steps. First, load scatter plot and convert it to a distance graph image. Process the graph image by applying a morphological Open operation to each connected graph - the steps of erosion and dilation constituting the open operation will alter node values. Next, for each node in turn, eliminate all edges whose weight exceeds the node's new value. Then, partition

15   the graph into connected components, a connected component or cluster being defined as a sub-graph of connected nodes - each node within a connected subgraph can be visited by traversing edges. Finally, filter out small groups and split large groups into two groups if necessary. The algorithm thus performs the following steps, as illustrated in Table VI. Table VI shows the pseudocode for the operation of

20   constructing a Decoding Map by fast cluster analysis.

49

Table VI

```
    CreateScatterPlot(Image, IntensityArray(FirstColor), IntensityArray(SecondColor)) ;
5   ConstructDistanceGraph(GraphImage, Image);   /** assign value to each node **/

    Open(GraphImage);

    For each(Node in the GraphImage){          /** process all nodes in the graph **/
10          For each(Edge of the Node){         /** process all edges of a node **/

                    If( Edge > Node){
                            DeleteEdge( );
                    }
15          }
    }

    PartitionGraph(GraghImage);                 /** partition graph into
                                                   connected subgraphs
20                                                 ("clusters") **/

    PostProcess(GraphImage);
```

Three-Color Encoded Objects: "Stacked" Decoding Maps. The clustering algorithm is applied to handle multi-dimensional cluster analysis for populations constructed as stacked two-dimensional clusters by preceding the clustering operation with a sorting step. In stacked two-dimensional clusters the third decoding image acquired in a case of encoding by three-color combinations will have one or more discrete intensity levels. Considering first the case of just a single intensity level, particles are readily sorted into two groups, namely those containing the third dye (labeled ON) and those not containing the third dye (labeled OFF) to obtain a stack of two-dimensional scatter plots which are individually analyzed to generate two corresponding decoding maps. In practice, this operation is performed using the

50

DECODER as follows. First, generate a two-dimensional scatter plot for the original

two dyes, encoding colors, designated G and B. This "G-B" plot represents an

intermediate result that corresponds to the projection of the three-dimensional "R-G-

B" space onto the "G-B" plane. To split this projected scatter plot into its constituent

5   components, generate a two-dimensional scatter plot just for third color, designated

R, in the "R-R" plane by providing two copies of the decoding image recorded in the

R- channel. The "R-R" plot will have the same size (and number of eventual

clusters) as the "G-B" plot (containing, after all, images of the very same objects).

The "G-B" plot is now split into two plots, one containing only points corresponding

10   to "R-OFF", the other containing only points corresponding to "R-ON". A (two-

dimensional) decoding map is now constructed for each plot using one of the

algorithms described above. This strategy is readily generalized to populations

encoded using multiple levels of a third color.

II.3.2   Combining Assay and Decoding Images

15   The identity of the binding agent of the binding agent-analyte complex is

determined by decoding. This step entails comparison of decoding image(s) and

assay image(s). That is, the assay image is sampled in accordance with the cluster

information in the decoding map to group assay signals by bead type and hence by

encoding tag as described above. A robust matching algorithm ensuring alignment of

20   decoding images and assay images is described below.

        Decoding may be carried out at the user site or at a central location. For

example, decoding images - in a suitable representation such as merged decoding

intensity arrays - are made available to the user, either in the form of text files on a

recording medium that is distributed along with bead arrays or in the form of a downloadable file available by way of authenticated access to a central database. Alternatively, decoding may be carried out on a central server after uploading of the assay image from the user site. Transaction protocols for this and related mode of data communication are disclosed below.

Matching by Variance Minimization (MVM). By construction, constituent beads of a randomly encoded bead array are randomly dispersed over the array. Assay signals recorded from a set of beads randomly drawn from all subpopulations or types of beads, these beads displaying different types of probes, will exhibit an inter-population variance, V, that reflects the differences in the corresponding probe-target molecular interactions. Assuming equal abundance of all bead types, this inter-population variance may be approximated by the variance associated with the distribution of the mean assay signals evaluated over each subpopulation, namely:

$$V = \sum_j (<<I>> - <I>_j)^2 ,$$

$<I>_j$ denoting the mean assay signal of the j-th subpopulation and $<<I>>$ denoting the mean of the $<I>_j$. In contrast, assay signals recorded from a set of beads drawn from the same subpopulation, these beads displaying the same type of probe, will exhibit an intra-subpopulation variance, v, that reflects aspects of characteristic remaining chemical heterogeneities such as bead size, density of probes displayed on beads, assay binding efficiency, etc, given that all probe-target interactions within

52

the subpopulation are nominally identical for each subpopulation; the intra-population variance has the form:

$$v = \sum_k (<I> - I_k)^2 ,$$

I_k denoting the assay signal recorded from the k-th bead within the subpopulation. Except in special circumstances, assay signals recorded from different subpopulations will be uncorrelated, and the variance, V, will exceed v:

$$V >> v$$

It is this insight which forms the basis for a robust "matching by variance minimization" algorithm by which to perform the cross correlation of decoding and assay images recorded from a random array and to resolve the task of perfectly aligning the two (or more) images of interest in the absence of fixed alignment aides and in the presence of edge-corrupting noise. That is, only in the correct alignment of decoding image and assay image are assay signals within the assay image sampled over members of the same subpopulation. In one embodiment, the alignment is performed by monitoring the variance, computed for one or more specific subpopulations of beads that may be included in the array for this purpose, as the assay image is shifted with respect to the decoding image.

As discussed herein in previous section II.2.2, even a single step displacement will completely scramble the sampling and mix assay signals from multiple subpopulations. Accordingly, the correct alignment, even in the presence of

53

considerable edge corruption, is robustly indicated by minimizing the variance of assay signals over a subpopulation as a function of relative displacement. In practice, "dark" particles or objects are filtered out during this matching step to eliminate erroneous contributions to the variance. The *FilterDarkBeads* function is required to

5   eliminate from the assay image objects or microparticles which do not contribute a measurable signal because - while the center of bright objects coincides with the maximum in the intensity profile it coincides with the minimum in the intensity profile for dark objects. This can lead to errors in aligning assays and decoding images.   In practice, each array is designed to include one or more reference

10   subpopulations displaying positive or negative control probes which are designed to generate a signal of known magnitude so as to ensure that indeed $v \ll V$. These one or more reference subpopulations are sampled to minimize the corresponding subpopulation variances, v, as a function of displacement from perfect alignment. That is, v is evaluated as the assay image grid is shifted relative to the decoding

15   image grid. If the condition $v \ll V$ is not satisfied, the algorithm produces a warning to indicate a possible problem with image quality and provides a choice to abandon further analysis.

As illustrated in Figs. 9A and 9B, there are seven possible misaligned positions if the search is confined to single row and single column misalignments,

20   namely: (-1, -1), (-1, 1), (-1, 0), (0, 0), (1, 0), (-1, 1), (1, 1); that is, to account for the stagger introduced in the image grid derived from a hexagonal lattice (910, 920), seven distinct positions are checked, as illustrated in Fig. 9C.

The *Minimal Variance Matching* algorithm operates on Decoding and Assay Data Records and evaluates the desired variance as a function of unit shifts applied to the two images represented in the respective data records. Table VII shows the pseudocode for the operation of combining decoding and assay images by way of Matching by Variance Minimization.

Table VII

```
LoadDecodingData(DecodingDataRecord, DecodingMap);
LoadAssayData(AssayDataRecord, IntensityArray);

MinVariance = -1000;
MinVarianceLocation = 0;

/** Check Variance Produced by 7 Possible Unit Displacements - see Text **/
For (i=0; i< 7; i++){
        ShiftGridPosition(i);
        FilterDarkBeads();
        Variance[i] = Merge(DecodingMap, IntensityArray);

        IF(Variance[i] < MinVariance){
                MinVariance = Variance[i];
                MinVarianceLocation = i;
        }
};

WriteAssayData(MinVarianceLocation, AssayDataRecord);        /** Save
Location **/
```

Correction for Multiple Scattering Effects. Unless suspended in specially selected density matching fluids which generally will be incompatible with bioanalytical assays of interest, polymeric, ceramic or other microparticles will exhibit strong scattering of visible light so that multiple scattering effects are readily observed in

55

planar assemblies and arrays of such particles. For example, fluorescence emitted by a microparticle within such a close-packed planar array is diffracted by its nearest neighbors and possibly its more distant neighbors, a source of potential error in overestimating assay signals. This effect is strongly dependent on interparticle distance and may be diminished or eliminated by appropriate array and substrate design as well as choice of illumination and collection optics.

In addition, the system herein also offers a method of correcting for the effects of multiple scattering on the intensity distribution recorded from subpopulations of beads within a planar array. As with the MVM algorithm, this method takes advantage of the random spatial distribution of different bead types throughput the array. Randomly placed "blank" beads, drawn from a subpopulation that is at most weakly fluorescent for purposes of encoding, serve as local "antennae" to establish a random sample of excess fluorescence produced by way of diffraction by the nearest neighbor configurations encountered within the array. To correct for the principal global effect of multiple scattering on assay signals recorded from a random encoded array, the variance of this excess fluorescence signal is subtracted from the intra-population variance of all subpopulations.

III   Transmission of Image Data Records

III.1   Identifying and Tracking Bead Arrays

As described in U.S. Patent Application Serial Nos. 10/238,439, and 10/365,993, entitled: "Encoded Random Arrays and Matrices" (the specification corresponding to WO 01/20593), incorporated herein by reference, and as further elaborated herein in Sections II.2.1 and II.2.2., each bead array generates its own

unique ID ("Intrinsic ChipID" or "ChipID") and can be identified. This ChipID may be physically or logically linked to a CarrierID. For example, multiple bead arrays may be mounted on a multichip carrier comprising a bar code which is capable of recording the indentity of each of the bead arrays. The CarrierID is tracked in the

5   course of producing bead arrays and also is tracked in the course of acquiring, processing and analyzing assay images using the system of the invention. The IntrinsicChipID may be linked to a CarrierID, and to a further assigned ChipID which may be appended to the CarrierID ("Appended ChipID") or may be otherwise physically linked to the CarrierID. Unless specifically indicated below, the ChipID

10   shall be understood to refer to Intrinsic ChipID or Appended ChipID.

Samples of interest for chemical analysis including samples collected from patients for clinical or other testing, also may be given a sample or patient ID, *e.g.*, in the form of a barcode which may be tracked along with the CarrierID using a barcode scanner. In addition, methods of intra-analyte molecular labeling have been

15   previously disclosed. For these purposes, the addition of unique molecular external labels or internal labels, for example in the form of a DNA fingerprints, may be considered. Such labels and associated methods have been described in U.S. Patent Serial No. 10/238,439. Information derived from the examination of these molecular labels may be entered into decoding and/or assay image records by the system herein

20   to minimize sample handling error and to facilitate the secure exchange of information, as elaborated herein below.

III.2 Transmission Protocols

In one embodiment, an assay image may be submitted for analysis by transmission over a network connection to a central location. Software available on centrally located servers, in one embodiment involving the ANALYZER described herein, completes the analysis and makes results of the analysis available to authorized users for retrieval. The analysis-server model provides protocols governing exchange of information in one or more two party transactions between one or more users and a central server where data is analyzed (Fig. 10a), and further provides protocols governing the exchange of information in three-party transactions, involving patient, to provider to a testing center where data is analyzed and assay results are recorded (Fig. 10b). A three-party transaction may also involve a recipient, a mediator and a provider of information.

This model offers several advantages to users as well as suppliers of molecular diagnostics, particularly when instrumentation distributed to field locations is easy to use and maintain, while the analysis of the data obtained using the instruments is complex. To the users' benefit, the requirement for designated staff with the training and expertise to install, master and operate analytical software is eliminated. Rapid turn-around of even advanced data analysis is ensured by access to rapid network connections even in remote locations at a doctor's office or patient site. Suppliers benefit from the reduction in cost associated with the logistics of providing extensive technical software support while providing high-speed analysis on dedicated server hardware. In general, the analysis service provider, equipment manufacturer and assay developer all may be distinct parties. Additional

58

parties may participate in certain transactions. For example, the manufacturer of the chips and arrays, and the analysis service provider, may not be the same party.

The system herein provides for an analysis server model invoking transaction protocols such as those elaborated below that ensure the secure exchange of private information created, for example, in genetic analysis -- an issue of wide concern. Additional services, such as advanced analysis in the form of binding pattern matching via database searches or data archiving, are readily integrated. In one embodiment, the analysis server model applies to assays producing data in the form of images and the analysis of interest relates to the analysis and archiving of images.

In one embodiment of such a transaction, the exchange of information may relate to the completion of analytical chemical, biochemical and diagnostic test, and the participating recipient, intermediary and provider of the information of interest (*e.g.,* personal genetic information) may correspond to patient, testing center and (data) analysis provider, respectively. Protocols are set forth for the secure creation, exchange and storage of information

Transaction Protocols

Pattern Matching via Access to "Fingerprint" Database. A pharmaceutical company researcher ("CLIENT") submits data recorded from an assay performed in an array format to probe the interaction between a set of immobilized proteins ("receptors") and a second set of proteins or ligands provided in a solution that is brought in contact with the receptor array. The data may be in the form of a decoded assay data record as described herein. Intensities - recorded from bead arrays or from other arrays, including "spotted" arrays - reflect a certain pattern of

interactions between receptors and ligands. Alternatively, the data may represent a pattern of expression levels for a set of designated genes of interest that may indicate an individual patient's response to treatment or may indicate a toxicology profile or may indicate the response triggered by a compound of interest that is to mimic the action of a known drug, said action being characterized on a molecular level by said expression pattern.

The two-party transaction between PROVIDER and CLIENT is performed in accordance with a protocol that preserves the anonymity of the CLIENT and simply permits the CLIENT to search a PROVIDER database for matching the interaction pattern or expression pattern with a unique pattern (or "fingerprint").

Advanced Services. Once decoded assay results are in hand, additional analysis may be optionally performed. In the simplest instance, statistical measures such as mean or variance are readily evaluated over each of the subsets. More generally, the presence of characteristic "patterns" of receptor-ligand interaction may be ascertained Such patterns may be indexed and stored in a searchable database to provide the basis for assay interpretation. This in turn will facilitate tracking and interpretation of disease histories and clinical trial results and aid in the identification of molecular identifiers and features ("genotype") associated with clinical pathology ("phenotype").

File Serving: Authenticated Remote Access to Decoding Image.- Decryption of the message contained within the assay image by application of the key represented by the decoding image groups intensities in accordance with bead type is provided. Following bead array assembly, the decoding image is analyzed to

derive a temporary ChipID representing a portion of the complete ChipID, based, for example, on the first row or column of the decoding image. The ChipID is stored, the temporary ChipID is transmitted to the user as a password for access to the database and retrieval of the full decoding image. In certain situations, it may be advantageous to the user not to download the full decoding image. For example, if assay results are negative, conclusions about a set of tested receptor-ligand interactions may be reached without decryption.

Alternatively, assay images may be uploaded to the server for additional analysis or archiving. Incoming assay images are linked to stored decoding images by matching temporary and full ChipID codes.

In this file server model, fees may be charged in accordance with the volume of transactions on a single transaction basis or on a subscription basis in accordance with pricing models practiced in the application server market.

A Two-Party Transaction Relating to the Analysis of Encrypted Assay Data.

The following two-party transaction (Fig. 10a) illustrated here using an embodiment in the form of custom bead arrays, invokes an encrypted covering to ensure that the identity of compounds in the assay remain private.

For example, a pharmaceutical company researcher ("CLIENT") provides to the custom bead array provider who, in this example, also is the analysis service provider ("PROVIDER"), a library of compounds to be subjected to on-chip assays in labeled containers with instructions to create an encrypted covering by simply recording the container labels corresponding to specific bead types. For example, "compound in container labeled A anchored to bead tag T1." The identity

61

of compounds within labeled containers is known only to CLIENT. In a preferred

embodiment, a unique set of bead tags is selected for a specific client to minimize

mishandling or inadvertent swapping of compound libraries.

The two-party transaction between PROVIDER and CLIENT is

performed in accordance with the following protocol:

Provider - Provide Bead Array with Unique ChipID

Create encrypted covering by attaching designated compounds to "tagged" beads

Create array encoding by assembling pooled beads into an array

Decode array configuration

Establish ChipID. In a preferred embodiment, the ChipID is derived from the array

configuration

Optionally, store ChipID in non-volatile memory to be packaged along with bead

array chip

Create a public database record ("Key") of the form (ChipID, Encrypted Covering)

Send bead array chip in assay cartridge (with ChipID) to Client

CLIENT - Perform Assay and Transmit Assay Image

Receive assay cartridge from Provider

Place analyte solution into assay cartridge and perform assay

Record ChipID. In one embodiment, use a chip carrier containing ChipID in

electronic representation in conjunction with an electronic reader and the array

imaging system such that the ChipID, read out from the assay chip, is recorded and

stored, and thus unambiguously linked, with the encrypted message in a public

record (ChipID/Public, Encrypted Message/Public)

Record assay image and create assay data record ("Encrypted Message")

Send combination of ChipID and Encrypted Message to PROVIDER for analysis

Provider - Perform Image/Data Analysis

Receive public record (ChipID, Encrypted Message) from CLIENT

5   Strip ChipID and check database for matching record (ChipID, Encrypted Covering)

Use ChipID to decode message, thereby creating a profile. In one embodiment, the

profile representing the decoded message has the form $\{<I>_k, 1 \leq k \ll n\}$, where the

$<I>_k$ represent intensities averaged over all beads of tag type k, tags being uniquely

associated with a specific compound in accordance with the encrypted covering

10   Create updated database record (ChipID, Encrypted Covering, Profile)

Provider - Transmit Profile

Supply database record (ChipID, Encrypted Covering, Profile) for retrieval by

CLIENT.

A Three-Party Transaction for the Secure Exchange of Genetic Information. By

15   generating, either concurrently with the completion of genetic analysis or by

concurrent analysis of tagging molecules added to patient samples, a molecular ID

such as a DNA fingerprint (as described in U.S. Patent Application Serial No.

10/238,439) that is embedded within the assay image, the methods of the present

invention create an unambiguous link between a chip ID ("IntrinsicChipID") derived

20   from the configuration of a random encoded bead array and a unique genetic ID,

thereby not only minimizing the possibility of error in sample handling but also

enabling verification of assay results and securing confidential genetic information in

the course of two-party or multi-party transactions, as elaborated below for a three-party transaction.

The process illustrated in Fig. 10C using a preferred embodiment in the form of custom bead arrays packaged on a carrier within a fluidic assay cartridge ensures the confidentiality of personal genetic information. Specifically, a three-party transaction between custom bead array provider who, in this example, also is the analysis service provider ("PROVIDER"), an intermediary or facilitator such as an assay service provider ("TESTTING CENTER") and user ("PATIENT") may be organized as follows to ensure privacy of information generated by genetic testing.

In this protocol, transactions between PROVIDER and TESTING CENTER and between TESTING CENTER and PATIENT involve public records that are identified by the ChipID. A separate transaction between PROVIDER and PATIENT involves the private information in the form of the genetic profile. The protocol below ensures that the identity of the PATIENT is concealed from PROVIDER: the PATIENT is identified only by the genetic ID presented for authentication in the final retrieval of genetic information. On the other hand, in the protocol below, genetic information is made available – by way of a "Relay" step - to the TESTING CENTER - or a designated physician or genetic counselor - for communication to the PATIENT.

The three-party transaction (Fig. 10C) between PROVIDER, TESTING CENTER and PATIENT is carried out in accordance with the following protocol.

Provider - Provide Encoded Bead Array Chip with Unique ChipID

Create covering by attaching probe molecules to color-encoded beads

64

Create array encoding by assembling pooled beads into an array

Decode configuration to establish ChipID

Optionally, store ChipID in non-volatile memory to be packaged along with bead array chip

5   Create a database record ("Key") of the form (ChipID/Public, Covering/Private)

Send packaged chip (with ChipID) to TESTING CENTER

PATIENT - Request Analysis

Submit sample to TESTING CENTER

Receive ChipID from TESTING CENTER

10   **TESTING CENTER - Perform Assay**

Receive assay cartridge from PROVIDER

Collect sample from PATIENT into assay cartridge

Send ChipID to PATIENT

Complete sample preparation and perform genetic analysis

15   **TESTING CENTER - Record Assay Image and Transmit Assay Data Record**

Record assay image with embedded GeneticID ("Encrypted Message")

Send combination of ChipID and Encrypted Message to PROVIDER for analysis. In a preferred embodiment, use a chip carrier containing ChipID in electronic representation in conjunction with an electronic reader and an image acquisition

20   system under general processor control such that the ChipID, read out from the assay chip, is recorded and stored, and thus unambiguously linked, with the encrypted message in a public record (ChipID/Public, Encrypted Message/Public)

For later verification: store assay cartridge, optionally containing patient blood

sample

Provider - Perform Image/Data Analysis

Receive public record (ChipID/Public, Encrypted Message/Public) from TESTING

CENTER

Strip ChipID and check database for matching decoding data record (ChipID/Public,

Covering/Private)

Use covering to fully decode message identifying genetic profile and embedded

GeneticID

In one embodiment, the decoded message has the form of intensities, $\{<I>_k, 1$

$\leq k \leq n\}$, averaged over beads of the same type, each type uniquely

identifying a specific probe molecule; other representations also are available

as discussed herein.

Extract GeneticID from decoded assay data record image as the subset of intensities

$<I>_k$ corresponding to ID-specific probes

Create updated database record (ChipID/Public, Covering/Private, GenID/Public,

GenProfile/Private)

Provider - Transmit Genetic ID

Send (ChipID/Public, GenID/Public) to Testing Center for transmission to Patient

PATIENT – Receive Genetic ID and Retrieve Genetic Profile (1080)

Using (ChipID/Public, GenID/Public), query Provider database

Provider - Transmit Genetic Profile (1090)

Authenticate GenID to authorize retrieval of private genetic profile from database

If, and only if, authentication confirmed, retrieve (ChipID/Public, Covering/Private, GenID/Public, GenProfile/Private)

Supply database record (GenID/Public, GenProfile/Private) for retrieval by

5   PATIENT

Using an assay cartridge, a physical linkage can be created between patient sample, assay cartridge and BeadChip with associated ChipID while the embedded genetic ID creates a physical linkage between genetic identity and genetic profile as an inherent part of the assay. Verification is then always possible by retesting. The

10   physical and logical linkages created by the methods of the present invention between patient sample, ChipID and genetic profile with embedded genetic ID eliminate common sources of error in genetic testing such as switching of patient samples.

Other transaction protocols may be devised using data structures of the type

15   introduced in the foregoing example, to ensure that only the PATIENT has access to genetic (or other) information created in an assay performed at the TESTING CENTER. For example, in one embodiment of the present invention, the PATIENT already may be in possession of his/her GeneticID prior to initiating a three-party transaction. In that case, the steps of transmitting, relaying and receiving GeneticID

20   (Fig. 10c, 1050, 1060, and 1070) may be eliminated. Instead, the PATIENT directly requests transmission of the genetic profile from the PROVIDER – access to the relevant database may be authenticated by comparing the Genetic ID used in the request with the Genetic ID extracted from the genetic profile.

Decoded assay data records may be archived. Archived decoded assay data records would be accessed only by those in possession of a GeneticID or equivalent key embedded in the decoded assay data record. That is, the database of archived records would be searched by rapid cross-correlation with the authentication code.

5   More generally, the following three-party protocol ensures that only the PATIENT (or his/her designee, such as a physician) who initiates a testing procedure is in possession of private information created in the test performed at the TESTING CENTER and analyzed by the application service PROVIDER. The TESTING CENTER has no access to the private information and PROVIDER has no

10   knowledge of the identity or particulars of the PATIENT. The PATIENT, having requested and having been assigned, a ChipID and SampleID, requests, directly from the PROVIDER, a confidential authentication key. In one embodiment, this is accomplished by access to the PROVIDER site, for example by a remote login. If the confidential authentication key generated by the combination of ChipID and

15   SampleID is taken by a third party, the PATIENT will have immediate knowledge that the confidential authentication key may be at risk of disclosure to a third party. The PATIENT may be able to request a new SampleID before providing a new biological sample to the TESTING CENTER. Software then assigns a randomly selected encrypted personalized authentication key to the combination of

20   ChipID/SampleID presented in the request. Only one such assignment is permitted. In one embodiment, the encrypted authentication key has the form of a "cookie" that is placed in a hidden directory on the hard drive or other storage device of the requesting machine so that only that machine is authenticated for future retrieval of

testing data from the PROVIDER. The PATIENT will ensure the integrity of the process: should an unauthorized party, for example, at the TESTING CENTER, attempt to acquire an authentication key, the subsequent attempt by the PATIENT to do so would fail, alarming the PATIENT to a possible breach in protocol. In one embodiment, the encrypted authentication key assigned to the requesting ChipID/Sample ID combination will be the IntrinsicChipID, or information embedded therein. In one embodiment, the random string of integers indicating vertex positions of a designated specific bead type within the BeadChip may be used. Following submission by the TESTING CENTER of the Assay Data Record, identified by a ChipID, for analysis, the PROVIDER combines the Assay Data Record with the Decoding Data Record corresponding to the submitted ChipID so as to create a decoded Assay Data Record from which specific embedded information such as a genetic profile may be extracted by "De-Covering", that is, application of the Covering to identify specific probes within the array as previously elaborated herein. This information is made available for retrieval by the PATIENT using the encrypted authentication key previously assigned to the ChipID/Sample ID combination. In one embodiment, only the machine previously endowed with a "cookie" will be permitted to access the database containing the requested information. This protocol ensures that the TESTING CENTER knows only the identity of the PATIENT but not the information such as a genetic profile extracted from the assay while the PROVIDER knows the information such as a genetic profile but not the identity of the PATIENT.

It will be apparent to those skilled in the art that the foregoing specific instances of two-party and three-party transactions merely illustrate the concepts involved which are applicable to a wider range of applications.

Pricing Strategies. The analysis server model of the present invention provides "fee-for-service" - in a single transaction format or in subscription pricing format - in which the initial cost of instrumentation as well as the recurring cost for disposable items can be absorbed in the charges for one or more of a palette of services. This has the advantage of eliminating user capital expenditures. The charges are for analysis, not for enabling instrumentation or assay components.

EXAMPLES

Example 1: Acquiring and Processing Decoding Image(s)

Fig. 11 illustrates the processing steps performed by the DECODER. Displayed in a Graphical User Interface (GUI) (1100) are three images, namely: a Brightfield image (1110), a green fluorescence image (1120) and a blue fluorescence image (1130). Grids, extracted and aligned by the DECODER, also are displayed. Also displayed is a scatter plot (1140) produced by the DECODER from the intensities in green and blue fluorescence images.

Example 2: Constructing Decoding Map

Fig 12 illustrates the construction of the 2D decoding map by the ANALYZER which may be integrated with the DECODER whose GUI is shown (1200). The map of the present example is composed of 33 clusters (1210), each of which is assigned a unique tag index. This is displayed for each cluster along with the number of beads contained in that cluster. For example, cluster 1 contains 101 beads.

70

Example 3: Acquiring and Processing Assay Image(s)

Fig. 13 illustrates the processing steps performed by the READER. Displayed in a GUI (1300) are three images, namely: a Brightfield image (1310), a fluorescence image (1320). Grids, extracted and aligned by the READER, also are displayed.

Example 4: Analyzing Images and Extracting Representations

Fig. 14 illustrates the Decoded Assay Data Record in two sections, namely: a text display (1400) listing assay signals extracted from the assay data record along with tag indices assigning each signal to a cluster and hence to a color code; and, a bar graph display (1420) of data from the Decoded Assay Data Record.

It should be understood that the foregoing examples and descriptions are exemplary only and not limiting, and that all methods and processes set forth are not to be limited to any particular order or sequence, unless specified, and that the scope of the invention is defined only in the claims which follow, and includes all equivalents of the claims.